IoT Architecture for Smart Cities Leveraging Machine Learning and SDN

Igor Miladinovic, Sigrid Schefer-Wenzl, and Heimo Hirner

Abstract—**Smart cities are expected to offer a variety of innovative digital services and to improve quality of life of their citizens. To achieve this, a huge number of sensors permanently monitors different objects generating vast amounts of data which need to be transmitted and processed efficiently. Today's networks struggle to scale and manage this data. Novel concepts, such a Software Defined Networking (SDN) and Multi-Access Edge Computing (MEC), are introduced to overcome these challenges. In this paper we enhance an IoT architecture for smart cities by integrating SDN for the optimal allocation of smart city applications between MEC centers and the cloud. This decision is taken using machine learning based on historical and current data. Some smart city applications illustrate this approach.**

Index Terms—**IoT, Machine Learning, Smart Cities, Multi-Access Edge Computing, SDN**

I. Introduction

Smart cities need an efficient communication infrastructure to cope with a rising number of digital services. An important enabler for these services are connected devices capable to communicate without human interaction, also known as Internet of Things (IoT) devices. Forecasts by Gartner [1], Nokia [2], or Ericsson [3] show all in the same direction—several tens of billions of IoT devices in the near future. To manage these devices and the associated growth in network traffic, today's communications networks need to evolve providing a new architecture capable to serve people and things optimally [4].

It is expected that IoT architectures for smart cities will increase the quality of public administration through continuous measurements of urban data and adaptation of the behaviour of people and things accordingly [5], [6]. Data from various sources, such as road sensors, cameras or vehicles, need to be collected, analyzed and evaluated in order to gain the current status of these resources and services, and to identify improvement potential. However, this data must be managed efficiently to avoid an explosion of network traffic.

Igor Miladinovic—Computer Science and Digital Communications, University of Applied Sciences Campus Vienna, Favoritenstrasse 226, A-1100 Vienna, Austria (phone: +43-1-6066877-2131; e-mail:igor.miladinovic@fh-campuswien.ac.at

Sigrid Schefer-Wenzl—Computer Science and Digital Communications, University of Applied Sciences Campus Vienna, Favoritenstrasse 226, A-1100 Vienna, Austria (phone: +43-1-6066877-2141; e-mail:sigrid.schefer-wenzl@fh-campuswien.ac.at

Heimo Hirner—Computer Science and Digital Communications, University of Applied Sciences Campus Vienna, Favoritenstrasse 226, A-1100 Vienna, Austria (phone: +43-1-6066877-2140; e-mail:heimo.hirner@fh-campuswien.ac.at

One of the emerging technologies designed to improve Quality of Experience (QoE) and to reduce network traffic is Multi-Access Edge Computing (MEC) [7]. By moving applications from the cloud near to end users, MEC is capable of serving time-critical applications, where a cloud solution—due to high latency—is not eligible. At the same time, MEC also reduces the network traffic, because the communication does not need to go completely in the cloud. As the resources in the MEC center are limited, the decision which applications are located in the MEC center directly influences both, QoE and network traffic. In addition, Software Defined Networking (SDN) represents a new approach to address management and scalability issues of today's networks [8]. SDN is a highly scalable and centralized network control architecture, which enables a global orchestration of distributed cloud and IoT resources.

In our previous work we have designed a layered IoT architecture for smart cities with focus on network traffic optimization by dynamically selecting the optimal application allocation between MEC centers and cloud. We introduce a Machine Learning (ML) module for the continuous monitoring of the potential traffic savings of each application, predicting the saving in the near future, and taking decisions which applications to move in the MEC center and when.

In this paper we extend this architecture by considering both, latency and network traffic, and including network conditions for this allocation. Network conditions are gathered from a SDN Controller over its northbound interfaces and utilized by the ML module for an optimal, dynamic allocation of applications. Depending on the network conditions, either applications generating higher traffic or having higher latency requirements are allocated in the MEC center. Therefore, when the network conditions allow for it, the architecture optimizes QoE by minimizing latency of relevant applications. Otherwise, if the network is heavily utilized, the architecture focuses on network traffic reductions.

II. Layered IoT Architecture for Smart Cities

One of the main challenges for smart city IoT solutions in today's infrastructure is the fragmented approach for operation, finance, regulation, and planning of city assets and services. Administration units often invest in independent solutions of a single problem, starting with a pilot project building up the infrastructure for that problem. This results in silos of infrastructure and IT resources [9]. A precondition for an efficient smart city is an open IoT architecture enabling all

978-1-7281-4790-1/19/$31.00 © 2019 IEEE

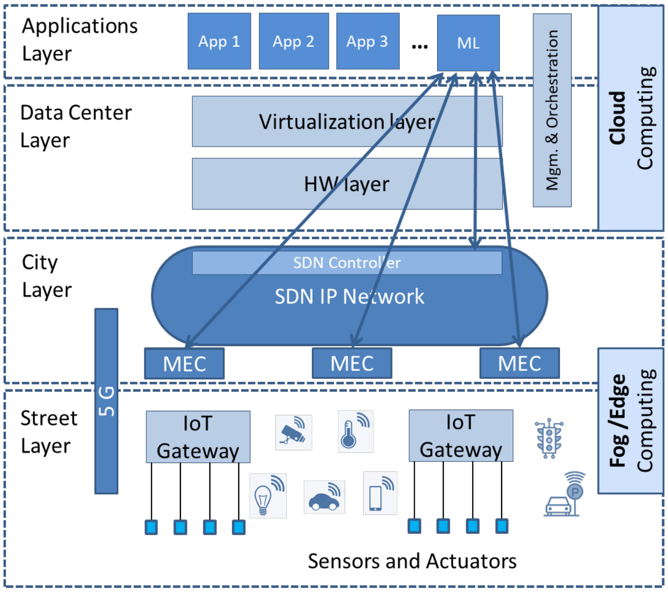

Fig. 1: Proposed IoT architecture

public services to use a common infrastructure exchanging data for cross-optimization.

Fig. 1 shows our proposed layered IoT architecture for smart cities, where some devices are connected directly with the network and some over IoT gateways. At the street layer, sensors, actors, devices and networking components are connected either directly or via an IoT gateway to the IP network. Examples of streets layer components are magnetic sensors, video cameras, lighting controllers, traffic lights, and connected cars. The city layer is composed of network routers and switches building an IP network dimensioned to manage the traffic amount of the city data that need to be transported. The centralized data center enables IoT applications to be running in the cloud on a standard hardware shared among different applications.

For the connection between the street layer and the city layer there is a need for a new, converged access architecture, capable of connecting a large number of devices efficiently [4]. The new generation of mobile networks, 5G, is expected to satisfy this need [10]. 5G focusses on eight major requirements [11], all highly relevant for IoT:

(1) up to 10 Gbps data rate,
(2) round trip latency of 1 ms,
(3) high bandwidth in unit area,
(4) very large number of connected devices,
(5) five-nines availability (99.999%),
(6) almost full coverage,
(7) up to 90% reduced energy consumption, and
(8) long battery live of connected devices (which is related to the 7^{th} requirement).

Therefore, we consider 5G as the primary access technology in our smart city architecture, as shown in Fig. 1. The devices do not need to be wired, the energy consumption is low (supporting also battery powered devices), the coverage will be given and time critical applications are supported.

A. Cloud vs Fog Computing

The cloud approach enables very high storage and processing capabilities, high scalability, automatic upgrades and updates of applications, potentially high security, and a centralized view for decision making. Besides these advantages, using the cloud for smart city applications results in some drawbacks [12]. The most critical one is the latency. The propagation time in optical fibre depends on the refractive index, which is different for different wavelengths. However, it is slightly below 1.5 resulting in a propagation time in optical fibre of approximately 2/3 of the light speed in vacuum (more exactly, it amounts to 204,357km/s for 1310nm wavelength). Therefore, a distance of 5,000km to a cloud data center causes a latency of approximately 45ms by the round trip time only ($rtt = 2 * \frac{5{,}000km}{204{,}357km/s}$). There are usually multiple hops to the cloud data center and each of them adds some processing time, resulting in a latency which is not acceptable for time critical applications. Moreover, as the processing time and the route are not constant, jitter is also high for applications running in the cloud. Other drawbacks of the cloud include increased network traffic, as the data first has to be transferred to the cloud and after processing back to the IoT devices in the street layer, and a high dependency on the network availability.

To address this issue, smaller data centers can be introduced and placed between the cloud and IoT devices, usually near the IoT devices. This decentralized computing infrastructure is referred to as fog computing [13]. However, these smaller data centers have limited resources, including computing power and storage capacity [9]. Furthermore, the complexity is increased, as applications running in several data centers need to be synchronised. Distributed architecture of fog computing is potentially more vulnerable, as any of data centers can be the target of a cyber-attack. Table I summarizes the benefits and drawback of cloud and fog computing.

TABLE I: Comparison cloud and fog computing

Parameter	Cloud	Fog
Latency	high	middle/low
Jitter	high	low
Network Traffic	high/middle	low
Computing power	high	middle/low
Storage capacity	high	low
Security	high	middle
Architecture Complexity	low	middle

B. Multi-Access Edge Computing

Inspired by the benefits of fog computing and aiming to improve the latency of applications in 5G networks, the European Telecommunications Standards Institute (ETSI) MEC Industry Specification Group (ISG) [14] specified an emerging technology called Multi-Access Edge Computing (MEC) [7]. The concept of MEC is to offer cloud computing functionalities within the Radio Access Network (RAN), very close to end customers and IoT devices. This reduces the latency and the network traffic compared with the cloud approach, and increases availability—as the same application can be deployed several times on different RANs [15]. As MEC is focused on computing close to end-devices outside the core of the network, it can be seen as a special case of fog computing.

Originally, MEC stood for Mobile Edge Computing, as it was first developed for mobile networks. As MEC development advanced, it was clear that it can be used with other access technologies. In order "to embrace the challenges in the second phase of work and better reflect non-cellular operators" ETSI changed its name from Mobile Edge Computing ISG to Multi-Access Edge Computing ISG.

With MEC, mobile network providers gain an important asset from the architectural point of view. During the last decades telecommunications network providers have been investing a lot in the networks to ensure connectivity, availability and quality of services. However, the majority of applications are cloud based and can be used independently of the network providers, making them easily replaceable by downgrading to connectivity providers only. With MEC, telecommunications network providers can enhance their customer value, as some of the applications will be running on their premises reducing latency to milliseconds and allowing for constant connectivity.

C. Software Defined Networking

Software Defined Networking (SDN) introduces a new architecture where the network control is separated from the data forwarding [16], [17]. The network is divided into three layers: (1) data plane layer, (2) control plane layer, and (3) application layer. The data plane layer is composed of programmable network infrastructure, such as switches and routers. The control plane layer comprises a centralised controller that manages the network devices based on requests from the application layer. Finally, the application layer is the place where networks services—fully implemented in software—are running and communicating with the centralized controller, the SDN controller.

Although SDN is a stand-alone technology, it can be combined with Network Function Virtualization (NFV) resulting in several synergy effects [18], [19]. For example, with the centralized network intelligence provided by SDN it is possible to dynamically move a network function from one data center to another without any interruption of service. Furthermore, network resources needed to meet some critical parameters (e.g. data rate or latency) can be allocated dynamically by SDN on request of a network function or an application. The main drivers for SDN and NFV are infrastructure agility and software-oriented innovations, resulting in higher scalability, costs savings, and shorter time to market of new services.

III. Application Allocation Considering Network Conditions

Smart city applications are very heterogeneous regarding the generated data traffic and latency requirements. Some applications are time-critical (e.g. traffic incident detection), some generate a lot of data (e.g. video surveillance), so that the selection of the applications located in the MEC center should be done carefully. On the one hand, MEC reduces the network traffic and enables a network latency of a few milliseconds [20], which is vital for time-critical applications. On the other side, MEC centers have limited resources, such as processing power and storage capacity. The question we are addressing in our research is which applications should be optimally located in MEC centers. Time-critical applications are always located in MEC centers, as they cannot fulfil their latency requirements if located in the cloud. For any other application we need to define what does *optimal* mean. Should we focus on the reduction of network traffic or on the improvement of user experience by reducing the latency.

In our previous work [21] we identified that—as the usage of applications changes depending on the daytime, day in the week or season—the selection of applications located in the MEC center needs to adapt continuously. In [22] we propose a machine learning module, which continuously monitors the savings of network traffic for each application when located in the MEC center. Based on these measurements and historical data, the module selects applications with the highest network traffic savings to be located in MEC centers.

In this paper we consider both, network traffic and latency for the optimal selection of the applications located in MEC centers. The current and predicted condition of the network is also important for this decision. An underutilized network would allow us to focus on the user experience and reduce the latency of relevant application—even if the price we have to pay for that is an increase in the network traffic. In a network close to congestion the reduction of network traffic is the priority.

To include the network condition in our decision for the application in the MEC center, ML module must be able to learn about the network. For this purpose we suggest an interface between the ML module and the SDN controller (see Section II-C), as shown in Fig. 1. As the SDN controller maintains a view over the complete network, and also provides northbound interfaces (APIs) for network applications, we use this interface for the connection with the ML module. The ML module not only learns about the current network condition, but also makes predictions about network condition in the near future.

For illustration purposes, let us consider several smart city applications. For each application we have estimated latency requirements and the expected network traffic. However, the network traffic is highly context-dependent (e.g. it depends on daytime or day in the week):

1) **Smart Parking**—providing information about parking spaces availability and booking possibility.
   Latency requirements: medium
   Network traffic: medium
2) **Structural conditions**—tracking of vibrations and material conditions in buildings, roads, bridges and historical buildings
   Latency requirements: low

Network traffic: low

3) **Noise Monitoring—**observation of noise level in problematic areas (close to motorways, night-life and centric zones).
Latency requirements: low
Network traffic: medium
4) **Traffic Information—**Measuring of vehicles and pedestrian traffic to optimize driving and walking routes.
Latency requirements: medium
Network traffic: medium
5) **Smart Lighting—**on-demand and weather adaptive street lightings
Latency requirements: medium
Network traffic: low
6) **Smart Waste Management—**Monitoring of fill level in trash cans to optimize the waste collection routes
Latency requirements: medium
Network traffic: low
7) **Smart Roads—**Real-time warning messages according to dangerous climate conditions and unexpected events like accidents
Latency requirements: high
Network traffic: low
8) **Smart Tourist Guide—**Augmented reality supported tourist guide with interactive information about tourist attractions
Latency requirements: high
Network traffic: high

Applications with high latency requirements (Applications 7 and 8) are always located in MEC centers. The remaining resources in MEC centers are used for other applications considering network conditions. In an underutilized network, applications with medium latency requirements (Applications 1, 4, 5, and 6) are prioritised to be located in MEC centers. On the other side, in a congested network applications with higher network traffic (Applications 1, 3, and 4) will be located in MEC centers. Consequently, Application 2 can be always placed in the could.

IV. Conclusions and Future Work

In this paper we demonstrated how SDN and machine learning can be combined to optimize allocation of smart city applications between MEC centers and the cloud. Given that this selection is dynamically changing, our approach is continuously learning and making predictions to adapt this selection. Taking the network conditions into account, this selection either focuses on improving QoE or on reducing the network traffic.

In our future work we will further investigate the benefits and constraints of our approach. At our university we are currently implementing the presented architecture. This will allow us to leverage the advantages of MEC and SDN. We will present the results for different smart city use cases.

References

[1] Gartner Inc., "Gartner Says 6.4 Billion Connected "Things" Will Be in Use in 2016, Up 30 Percent From 2015," http://www.gartner.com/newsroom/id/3165317, Nov. 2015.

[2] Nokia, "A buyer's and influencer's guide to connected device management, for all things connected in mobile, home and IoT," Strategic Whitepaper, Jul 2016.

[3] Ericsson, "Ericsson Mobility Report, On the Pulse of the Networked Socienty," Report, Jun 2016.

[4] M. K. Weldon, *The Future X Network, A Bell Labs Perspective*. CRC Press, 2016.

[5] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, Feb 2014.

[6] E. Ahmed, I. Yaqoob, A. Gani, M. Imran, and M. Guizani, "Internet-of-things-based smart environments: state of the art, taxonomy, and open research challenges," *IEEE Wireless Communications*, vol. 23, no. 5, pp. 10–16, October 2016.

[7] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.

[8] P. Goransson, C. Black, and T. Culver, *Software Defined Networks: A Comprehensive Approach*, 2nd ed. Morgan Kaufmann, 2017.

[9] D. Hanes, G. Salgueiro, P. Grossetete, R. Barton, and J. Henry, *IoT Fundamentals: Networking Technologies, Protocols, and Use Cases for the Internet of Things*, ser. Fundamentals. Pearson Education, 2017.

[10] GSMA Intelligence, "Understanding 5G: Perspectives on future technological advancements in mobile," White Paper, Dec. 2015.

[11] M. Agiwal, A. Roy, and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 1617–1655, thirdquarter 2016.

[12] I. Miladinovic and S. Schefer-Wenzl, "A highly scalable iot architecture through network function virtualization," *Open Journal of Internet Of Things (OJIOT)*, vol. 3, no. 1, pp. 127–135, 2017, special Issue: Proceedings of the International Workshop on Very Large Internet of Things (VLIoT 2017) in conjunction with the VLDB 2017 Conference in Munich, Germany. [Online]. Available: http://nbn-resolving.de/urn:nbn:de:101:1-2017080613543

[13] M. Chiang and T. Zhang, "Fog and IoT: An Overview of Research Opportunities," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 854–864, Dec 2016.

[14] MEC ETSI ISG, "Mobile Edge Computing (MEC); Framework and Reference Architecture," ETSI GS MEC 003 V1.1.1 (2016-03), 2016.

[15] D. Sabella, A. Vaillant, P. Kuure, U. Rauschenbach, and F. Giust, "Mobile-Edge Computing Architecture: The role of MEC in the Internet of Things," *IEEE Consumer Electronics Magazine*, vol. 5, no. 4, pp. 84–91, Oct 2016.

[16] Y. Li and M. Chen, "Software-Defined Network Function Virtualization: A Survey," *IEEE Access*, vol. 3, pp. 2542–2553, 2015.

[17] T. D. Nadeau and K. Gray, *SDN: Software Defined Networks: An Authoritative Review of Network Programmability Technologies*. O'Reilly Media, 2013.

[18] S. Aleksic and I. Miladinovic, "Network virtualization: Paving the way to carrier clouds," in *2014 16th International Telecommunications Network Strategy and Planning Symposium (Networks)*, Sept 2014, pp. 1–6.

[19] J. Costa-Requena, J. L. Santos, V. F. Guasch, K. Ahokas, G. Premsankar, S. Luukkainen, O. L. Pérez, M. U. Itzazelaia, I. Ahmad, M. Liyanage, M. Ylianttila, and E. M. de Oca, "SDN and NFV integration in generalized mobile network architecture," in *2015 European Conference on Networks and Communications (EuCNC)*, June 2015, pp. 154–158.

[20] J. Zhang, W. Xie, F. Yang, and Q. Bi, "Mobile edge computing and field trial results for 5g low latency scenario," *China Communications*, vol. 13, no. 2, pp. 174–182, Nov. 2016.

[21] I. Miladinovic and S. Schefer-Wenzl, "Dymanic Allocation of Smart City Applications," *Open Journal of Internet Of Things (OJIOT)*, vol. 4, no. 1, 2018, special Issue: Proceedings of the International Workshop on Very Large Internet of Things (VLIoT 2018) in conjunction with the VLDB 2018 Conference in Rio de Janeiro, Brazil.

[22] ——, "Towards Intelligent Multi-Access Edge Computing Using Machine Learning," in *International Conference on Interactive Mobile Communication Technologies and Learning*, 2019.